



Syntheticus[®]

Deloitte.

*Reshaping Analytics
with Syntheticus[®]:*

A Solution for Data Scarcity and Limited Availability



Content

Introduction	2
What Is Synthetic Data?	2
How Is Synthetic Data Generated?	3
Beyond Data Privacy: The Advantages of Synthetic Financial Data	4
Deloitte's Analytics Success Story: Synthesizing a Relational Database	5
Best Practices and Considerations for Using Synthetic Data	9
The Future of Data-Driven Financial Services	11

Introduction

In the world of analytics, data is king. The more data you have, the more insights you can gain and the better decisions you can make. But what if you don't have enough data, or the data you have is biased, incomplete, or outdated?

That's where synthetic data comes in.

Synthetic data is artificial data generated by computer algorithms to mimic the patterns and characteristics of real-world data. Think of it as "digital siblings" of your data but with more flexibility and diversity. By using synthetic data, you can supplement or replace your real-world data, get more accurate and comprehensive results, and expand your analytical capabilities. And that's not just a theory.

Synthetic data has already proven valuable in various analytics domains and industries, from healthcare and finance to marketing and cybersecurity. In this case study, we will explore the concept of synthetic data and its potential to address the problem of limited data availability. By the end, you'll understand how synthetic data can help you make better-informed analytics decisions and drive your business forward.

What Is Synthetic Data?

At its core, synthetic data is a computer-generated representation of real-world data. But unlike real-world data, which is collected from various sources such as surveys, sensors, or databases, synthetic data is created from scratch based on predefined rules or statistical models. This gives it some distinct advantages over traditional data sources, such as flexibility, scalability, and privacy.

Synthetic data aims to provide additional data points similar to real-world data, allowing for more comprehensive analysis and modeling. It supplements existing data sets or creates new ones, enabling a more detailed look at trends and patterns.



Gartner

[Gartner](#) named "Synthetic Data" and "Differential Privacy" as one of its Top Strategic Technology Trends and estimates that 60% of large enterprises will be leveraging one or more of these techniques by 2025.



Forbes

[Forbes](#) named 'Synthetic Data' as one of the Top 10 transformative CV Trends in 2024, further highlighting its growing importance.

Since synthetic data isn't considered "Personal Identifiable Information (PII)," according to privacy regulations such as GDPR, all the data can be safely used and collaborated on without worrying about privacy breaches or compliance.



According to the European [Data Protection Supervisor \(EDPS\)](#), "Synthetic data is a technical solution to a legal problem," enhancing technology privacy, mitigating bias, and democratizing access to data. In their regularly published [TechSonar report](#) on emerging technologies, EDPS mentions synthetic data as one of the most promising technologies worth monitoring.

With growing concerns over data privacy and compliance, synthetic data has become a crucial tool for businesses and organizations looking to comply with regulations while still leveraging the power of data analytics. It helps bridge the gap between legal requirements and technological advancements, making it possible to use sensitive data for analysis without compromising privacy.

How Is Synthetic Data Generated?

Synthetic data is created artificially using Generative AI techniques, such as Generative Adversarial Networks (GANs) and Variational Autoencoders (VAEs), in combination with Differential Privacy. These techniques create synthetic data that realistically mimics the original data while preserving privacy.



Once the original data is collected, the platform orchestrates the proprietary machine learning algorithms to synthesize, secure, validate, augment, and enrich it. This results in synthetic data that looks and behaves like the original one without revealing PII or risking compliance issues.

Beyond Data Privacy: The Advantages of Synthetic Financial Data

When you think about synthetic data, the first things that come to your mind are probably data privacy and security concerns. While data privacy is paramount, synthetic data is much more than a privacy shield.

Generating synthetic data opens up a world of possibilities and improvements to your analytics, and many of those have nothing to do with privacy or security.

Here are some of the main benefits synthetic data will bring to your organization:

● Improved Data Quality and Diversity

Synthetic data is generated to represent a wide range of scenarios and events, allowing you to create more diverse and representative datasets than those available from traditional sources. This diversity improves the accuracy of your models and enables better predictions and risk assessments.

● Scalability Without Constraints

Traditional data sources often impose limitations on the scope and volume of data. Synthetic data, on the other hand, allows you to generate data on demand to support your ML algorithms. This scalability empowers you to expand your analytics operations without hindrance.

● Advancing Risk Management

Synthetic data simulates different risk scenarios, allowing you to test and refine your risk management strategies before implementing them in the real world. By running simulations on synthetic data, you will better understand and anticipate risk and minimize losses while reducing the time and resources needed to develop your models.

● Enabling CSP Analytics Tools

Using cloud service provider (CSP) analytics tools is a game-changer in the modern analytics landscape. However, data privacy concerns often limit their adoption. Synthetic data provides a secure path, enabling organizations to harness the power of CSP analytics tools without compromising data integrity or compliance.

● Improving Collaboration and Knowledge Sharing

Due to its privacy-preserving capabilities, synthetic data is easily shared and distributed within and between organizations, enabling better collaboration between departments and teams. By collaborating on synthetic data, you will build on your organization's collective knowledge and expertise, improving the quality of your models and predictions.

● Streamlining Regulatory Compliance

Synthetic data helps organizations comply with strict data privacy and security regulations, such as GDPR, CCPA, and the Swiss nFADP. It allows organizations to train their models without using real-world data, which may contain sensitive or personally identifiable information. This means you can test and validate your models while remaining compliant with regulations and avoiding potential legal liabilities.

● Reducing Storage Costs

Traditional data storage methods are expensive, especially for large datasets. However, with synthetic data, organizations significantly reduce their storage costs. Synthetic data is generated on demand and does not require physical storage space, making it a cost-effective solution for companies looking to scale up their analytics capabilities without breaking the bank.

● Fueling Innovation and Experimentation

Synthetic data empowers innovation and experimentation. It enables you to prototype and test new analytical approaches, products, services, and business models without risking real-world assets. You can experiment with confidence, knowing you can validate your ideas in a safe and efficient manner.

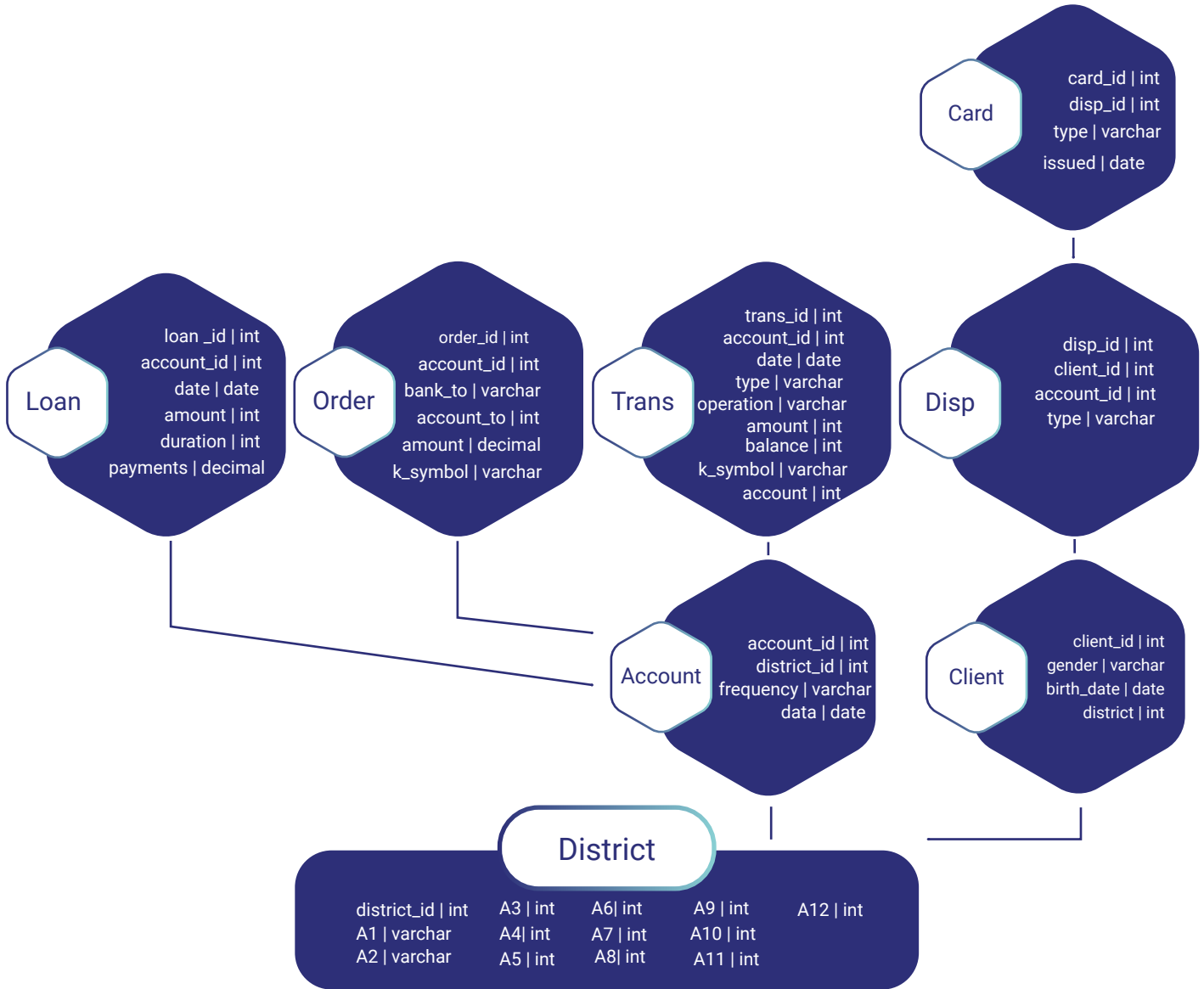
Deloitte. Analytics Success Story: Synthesizing a Relational Database

At Syntheticus®, we love tackling challenging projects that push the boundaries of what's possible with data and take pride in our ability to solve complex data challenges for our clients. In this spirit, we have formed a partnership with Deloitte to help them find a solution for their large and complex financial data.

They were struggling to manage **>8 million rows of data** spread across a dozen tables connected through referential integrity.



*Given the client's privacy, we're not disclosing the actual underlying data but only the schema of a public dataset for the sake of showcasing the results. See below for the entity relationship diagram.



To tackle the challenge, we leveraged our expertise in data modeling and database design to optimize the tables for performance and scalability while maintaining referential integrity. Using Syntheticus®, we seamlessly connected to the client's database. The out-of-the-box integrations support various services such as AWS, Google, Azure, Snowflake, Databricks, and integration with enterprise databases such as Oracle, MySQL, MongoDB, PostgreSQL, MariaDB, and many others.

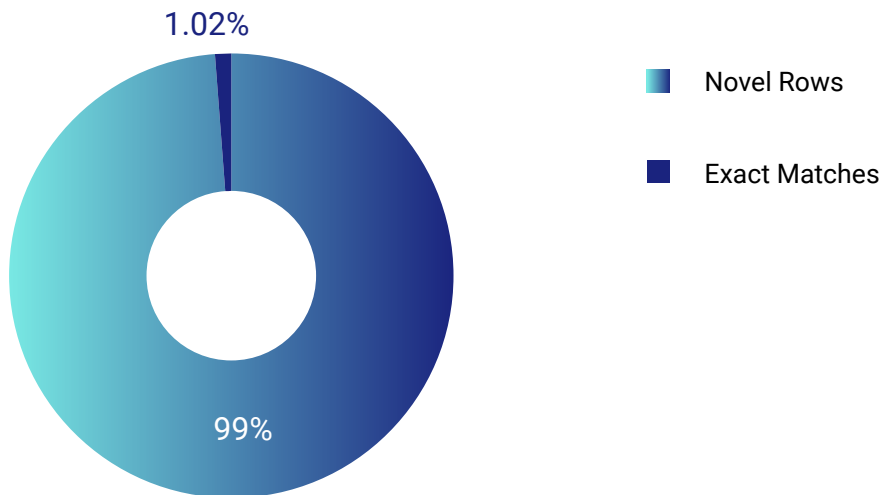


After only a few minutes of reviewing the client’s underlying data model, Syntheticus® generated a synthetic version of the entire database with a dozen tables, each with a specific set of fields to accommodate the client’s requirements, and over 8 million records.

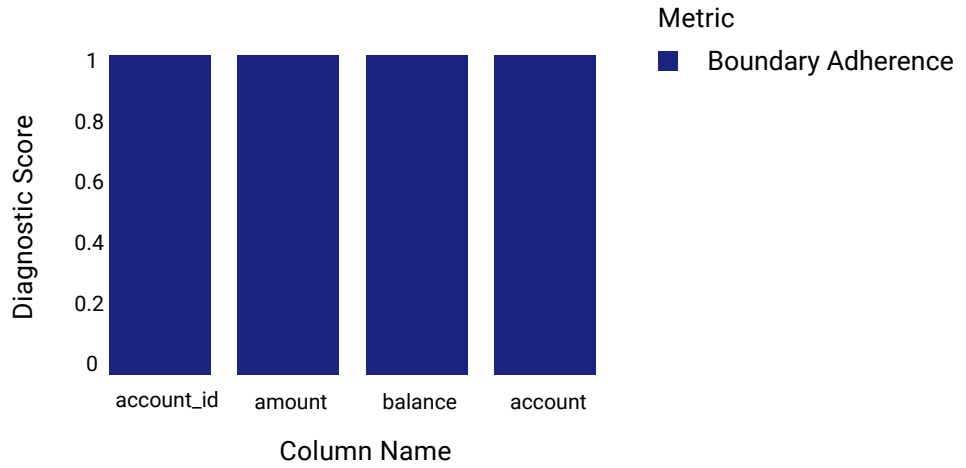
We take quality seriously, and as part of our commitment, the platform conducted a comprehensive Quality Assurance (QA) report. This report not only ensured that the database met the highest standards but also provided the necessary information for a Data Protection Impact Assessment (DPIA). The DPIA allows legal experts to define the synthetic data as anonymized, ensuring GDPR compliance. Additionally, the QA report highlighted the main metrics, including data utility, fidelity, and privacy for each single table as well as for the database.

Here are some of the main metrics found within the QA report:

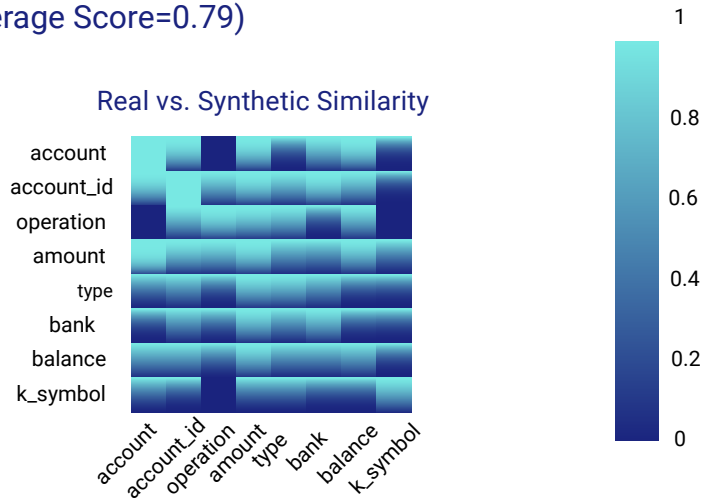
**Data Diagnostic:
Synthesis (Score=0.99)**



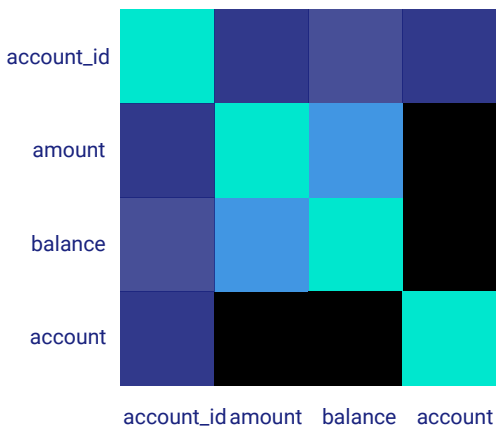
Data Diagnostics:
Column Boundaries (Average Score=1.0)



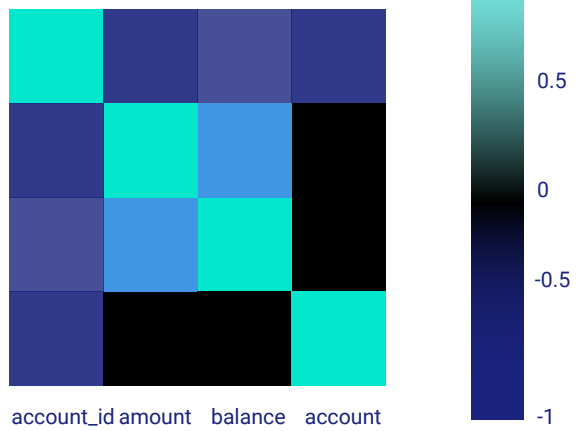
Data Quality:
Column Pair Trends (Average Score=0.79)



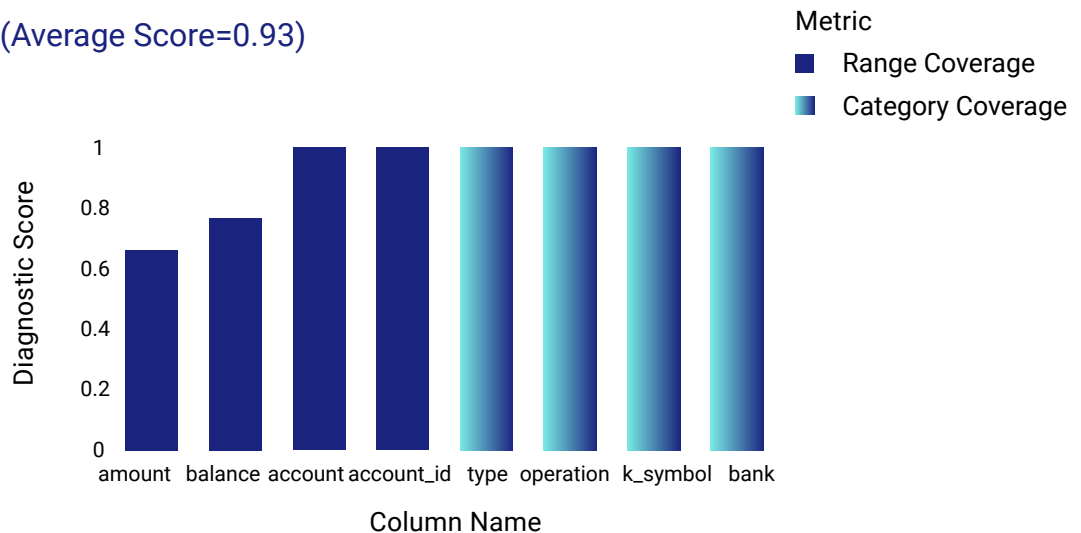
Numerical Correlation (Real Data)



Numerical Correlation (Synthetic Data)



Data Diagnostics: Column Coverage (Average Score=0.93)



After delivering the optimized and quality-assured relational database to Deloitte, we provided them with a set of recommendations to further improve the database's functionality and usability.

Our team at Syntheticus® takes pride in delivering solutions that meet and exceed our client's expectations and we're thrilled to have provided our client with a sophisticated database solution that simplifies their data management and supports their business goals.

Deloitte uses the synthetic data provided to bolster their analytical capabilities and to refine their Business Intelligence (BI) strategies. The high-quality, GDPR-compliant synthetic database acts as a robust foundation for their advanced data analytics.

Deloitte analysts can now conduct comprehensive data mining and predictive modeling without compromising the privacy of real-world individuals. The synthetic data's accuracy empowers Deloitte to derive meaningful insights, identify trends, and make informed strategic decisions, enhancing their competitive advantage in the market.

Best Practices and Considerations for Using Synthetic Data

If you've decided to leverage synthetic data for your institution, there are a few things to keep in mind to ensure you're using it effectively and responsibly.

First, it's crucial to select appropriate models that are relevant to your use case. Different models have different strengths and weaknesses, and you should choose one that fits your specific needs. For example, if you're looking to generate data for credit risk analysis, you might choose a model that accurately simulates default probabilities and other credit risk metrics.

But selecting the suitable model is just the first step. You need to validate the quality and reliability of the synthetic data you generate. This means running tests to ensure your data is accurate, unbiased, and representative of real-world data. Only then can you have confidence in the insights you extract from it.

Of course, ethical and legal considerations also come into play when working with synthetic data. Data privacy is a big one, as synthetic data is sometimes derived from real-world data containing personally identifiable information (PII).



Intellectual property is another concern, as some synthetic data generation techniques may involve proprietary algorithms or models.

Another important consideration when working with synthetic data is ensuring compliance with regulatory requirements. Depending on your industry and location, specific regulations or guidelines may dictate how to generate and use synthetic data. Make sure to do your research and consult with legal experts to ensure you're following all applicable regulations and guidelines.

Finally, the lack of clear standards on privacy can create uncertainty around how to best protect sensitive information in synthetic datasets. Syntheticus® recognizes this challenge and actively participates in the IEEE Standards Association, which has set up an IC Expert Group to set a standard for structured privacy-preserving synthetic data.

Transparency, explainability, and accountability are essential principles to keep in mind when using synthetic data.



You need to be able to explain how the data was generated, what assumptions were made, and what the limitations of the data are. This is important both for internal decision-making and for external stakeholders who may be relying on it.

The Future of Data-Driven Services

The analytics landscape is experiencing a significant shift driven by technological advancements and the critical role of data in decision-making. The future of analytics will be molded by organizations' ability to harness data-driven insights and successfully navigate the digital transformation. A pivotal challenge in this context is the effective handling and processing of data, including personal data, particularly within the realm of analytics and artificial intelligence.

Synthetic data is likely to become even more important for institutions in the years to come. The increased availability of synthetic data will allow institutions to leverage AI and machine learning to make more informed decisions while complying with data privacy and regulatory requirements.

As new technologies like blockchain and distributed ledgers become widespread, the potential applications for synthetic data will only grow. From risk management and compliance to fraud prevention, synthetic financial data will become an essential tool for analytics-driven enterprises.

However, this future is not without its challenges, including data privacy concerns, cybersecurity risks, and regulatory compliance. As the analytics landscape continues to evolve, organizations must strike a delicate balance between capitalizing on data-driven insights and mitigating potential risks to ensure sustainable success. Synthetic data will be essential to this equation, bridging the gap between traditional analytics and the data-driven future.



Syntheticus[®]

Ready to explore the power of **synthetic data** for your analytics projects?

[Sign up for a free demo](#)

and learn how Syntheticus[®] advances your
data-driven analytics projects while
protecting customer privacy.

syntheticus.ai